

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências exatas e de tecnologia

Departamento de Computação

Trabalho de conclusão de curso - TCC

Paulo Vitor Tognolo de Lima

**Clusterização de Dados no Mercado de Crédito
Privado: A Influência das Passagens entre
Fundos de Investimento**

São Carlos - São Paulo

2024

Paulo Vitor Tognolo de Lima

Clusterização de Dados no Mercado de Crédito Privado: A Influência das Passagens entre Fundos de Investimento

Trabalho de Conclusão de Curso apresentado
ao Departamento de Computação como parte
dos requisitos para a conclusão da graduação
em Engenharia de Computação.

Orientação Prof. Dr. Alexandre Levada

São Carlos - São Paulo

2024

Aos meus queridos pais que tiveram a força e a paciência de não somente trilhar a base para minha caminhada, como também me esperar vivê-la. Assistindo empolgadamente a construção do meu triunfo.

Agradecimentos

Agradeço a meu pai, Olinto por me ensinar a persistência através do exemplo do seu caráter e a minha mãe, Elenir que com seu coração amoroso e canetas coloridas conseguiu despertar em mim o anseio pelo conhecimento.

Agradeço também à minha companheira Alice, por ser o alicerce necessário na correria atribulada que a rotina paulistana proporciona enquanto não perde seu carinho.

Por último gostaria de agradecer ao meu amigo Lucas e meus companheiros de trabalho, por me apresentarem o mercado financeiro e seu mundo repleto de oportunidades.

*“É preciso desfrutar o simples prazer da execução de uma tarefa bem feita
(Zima Blue)”*

Resumo

Este trabalho tem como objetivo principal analisar o impacto das operações de *Passagens entre Fundos* no mercado secundário de renda fixa para crédito privado, com foco na debênture PEJA11. Para isso, foi desenvolvido um algoritmo de reconhecimento de padrões utilizando o método de aprendizado de máquina não supervisionado DBSCAN, com o intuito de identificar e diferenciar essas operações das negociações efetivas. A partir de uma análise detalhada dos dados de negociação registrados na CETIP, abrangendo o período de abril de 2023 a abril de 2024, foi possível observar como as *Passagens Entre Fundos* influenciam a métrica de *Average Daily Trading Volume* (ADTV) e a percepção de liquidez no mercado. A escolha da debênture PEJA11 foi motivada por seu alto volume financeiro e pela frequência significativa de suas negociações, tornando-a representativa para o estudo. Os resultados obtidos demonstram que o algoritmo desenvolvido foi eficaz na identificação de operações de *Passagem Entre Fundos*, resultando em uma redução de aproximadamente 28% no *Average Daily Trading Volume* (ADTV) após a aplicação do algoritmo, comparado ao ADTV calculado sem a utilização do método. Essa redução destaca a influência significativa das *Passagens Entre Fundos* na percepção de liquidez. Além disso, as técnicas de visualização de dados aplicadas forneceram insights claros e intuitivos sobre os padrões de negociação, facilitando a interpretação dos resultados. Este estudo contribui para uma melhor compreensão das nuances do mercado de crédito privado e sugere que a metodologia desenvolvida pode ser aplicada a outros ativos, auxiliando na tomada de decisões informadas e na precificação mais precisa desses títulos.

Palavras-chave: Mercado de Crédito Privado, Debêntures, Passagem Entre Fundos, DBSCAN, Volume Médio Diário de Negociação (ADTV), Liquidez, Precificação.

Abstract

The main objective of this work is to analyze the impact of *Transfer Procedures* operations in the secondary market of fixed income for local corporate bonds, focusing on the PEJA11 debenture. For this purpose, a pattern recognition algorithm was developed using the unsupervised machine learning method DBSCAN, with the aim of identifying and differentiating these operations from effective trades. Based on a detailed analysis of the trading data registered in CETIP, covering the period from April 2023 to April 2024, it was possible to observe how *Transfer Procedures* operations influence the *Average Daily Trading Volume* (ADTV) metric and the perception of liquidity in the market. The choice of the PEJA11 debenture was motivated by its high financial volume and the significant frequency of its trades, making it representative for the study. The results obtained demonstrate that the developed algorithm was effective in identifying *Transfer Procedures* operations, resulting in a reduction of approximately 28% in the *Average Daily Trading Volume* (ADTV) after the algorithm was applied, compared to the ADTV calculated without using the method. This reduction highlights the significant influence of *Transfer Procedures* on the perception of liquidity. Additionally, the data visualization techniques applied provided clear and intuitive insights into trading patterns, facilitating the interpretation of the results. This study contributes to a better understanding of the nuances of the local corporate bonds market and suggests that the developed methodology can be applied to other assets, assisting in making informed decisions and more accurate pricing of these securities.

Keywords: Local Corporate Bonds, Debentures, *Transfer Procedures*, DBSCAN, Average Daily Trading Volume (ADTV), Liquidity, Pricing.

Lista de ilustrações

Figura 1 – Os 10 ativos mais negociados no período.	25
Figura 2 – As 10 maiores movimentações financeiras de abril de 2023 a abril de 2024.	25
Figura 3 – Análise dos dados para a escolha do valor <i>eps</i> com base nos 12 clusters com maior número de operações.	27
Figura 4 – Aplicação do DBSCAN nas negociações do ativo PEJA11 com base na taxa de negócio.	28
Figura 5 – Distribuição da Contagem de Negócios por Clusters que não são Outliers.	28
Figura 6 – Refinamento dos clusters após a aplicação do DBSCAN, com remoção de falsos positivos.	29
Figura 7 – Análise comparativa dos resultados antes e após a aplicação do Algoritmo.	30
Figura 8 – Comparação das Taxas de Negociação Intradiárias antes e depois da Aplicação do Algoritmo.	31

Lista de tabelas

Tabela 1 – Comparação das Métricas de Liquidez antes e depois da Aplicação do Algoritmo	30
---	----

Sumário

1	INTRODUÇÃO	11
2	OBJETIVOS	14
2.1	Objetivos Gerais	14
2.2	Objetivos Específicos	14
2.3	Resultados Esperados	15
3	CONTEXTO ECONÔMICO E TEORIA DE MERCADO	16
3.1	Contexto Econômico	16
3.1.1	Títulos do Mercado de Crédito Privado	16
3.1.2	Indexadores Relacionados às Debêntures	17
3.1.3	Debêntures e Métodos de Precificação	17
3.1.4	Método de Precificação da ANBIMA	18
3.2	Importância das Medidas de Liquidez no Mercado Secundário	18
4	ALGORITMOS E TÉCNICAS RELACIONADAS	20
4.1	Escolha do Algoritmo Utilizado	20
4.1.1	Motivação para o Uso do DBSCAN	20
4.1.2	Considerações sobre a Não Utilização de Técnicas Supervisionadas	21
4.2	Uso do DBSCAN e Escolha do Parâmetro <i>Eps</i>	21
4.2.1	Escolha do Parâmetro <i>Eps</i>	21
4.2.2	Justificativa para a Escolha de <i>minPts</i> igual a 2	22
4.2.3	Adaptação do DBSCAN às Características dos Dados	22
5	METODOLOGIA DESENVOLVIDA E EXPERIMENTOS	23
5.1	Métodos	23
5.1.1	Extração dos Dados	24
5.1.2	Tratamento dos Dados	24
5.1.3	Separação de Debêntures e Seleção do Ativo PEJA11	25
5.1.4	Agrupamento por Data de Liquidação e Quantidade	25
5.1.5	Aplicação do Algoritmo DBSCAN	26
5.1.6	Remoção de Falsos Positivos e Refinamento dos Clusters	29
5.2	Resultados Experimentais	29
5.2.1	Análise Comparativa das Métricas de Liquidez	30
5.2.2	Influência das <i>Passagens entre Fundos</i> nas Taxas de Negociação Intradiárias	30
5.3	Discussão dos Resultados	31

6	CONCLUSÕES FINAIS	33
6.1	Trabalhos Futuros	33
	REFERÊNCIAS	35

1 Introdução

O mercado de crédito privado desempenha um papel fundamental no financiamento de empresas e projetos em economias ao redor do mundo. Esse mercado engloba uma variedade de instrumentos financeiros, como debêntures, certificados de recebíveis imobiliários (CRI), certificados de recebíveis do agronegócio (CRA), letras financeiras, certificados de depósito bancário (CDB), entre outros. Esses instrumentos são utilizados por empresas e outras entidades para captar recursos junto a investidores, oferecendo, em troca, títulos de dívida que prometem um retorno fixo ou variável.

No Brasil, o mercado de crédito privado tem crescido significativamente nas últimas décadas, impulsionado pela busca de empresas por alternativas de financiamento além das tradicionais linhas de crédito bancário. Esse crescimento tem sido facilitado por um ambiente regulatório que promove a diversificação de fontes de capital e incentiva a participação de investidores institucionais e individuais.

Dentro desse cenário, é essencial compreender as dinâmicas dos mercados primário e secundário de crédito privado. O mercado primário é o espaço onde os títulos são emitidos pela primeira vez e vendidos diretamente aos investidores, geralmente em operações conduzidas por instituições financeiras que atuam como intermediárias. Já o mercado secundário refere-se à negociação desses títulos entre investidores após a emissão inicial, permitindo a liquidez dos ativos e a gestão ativa de portfólios.

Mercado Primário: Envolve a negociação de títulos privados entre o emissor da dívida e os potenciais compradores. Este é o momento em que as empresas captam recursos para financiar seus projetos, estabelecendo as condições da dívida, como taxa de retorno, prazo de vencimento e garantias associadas.

Mercado Secundário: Refere-se à negociação de títulos privados entre um investidor (que já possui o título em carteira) e outro potencial comprador, geralmente por meio de corretoras certificadas. O mercado secundário é crucial para a liquidez dos ativos de crédito privado, permitindo que os investidores ajustem suas posições de acordo com suas estratégias e necessidades de liquidez.

No Brasil, todas as negociações de títulos privados no mercado secundário são registradas na Central de Custódia e Liquidação Financeira de Títulos Privados (CETIP), que agora faz parte do conglomerado da B3 (Brasil, Bolsa, Balcão). A CETIP disponibiliza o histórico das últimas 20 sessões de negociação, oferecendo transparência e informações cruciais para o mercado.

Dentro do mercado de crédito privado, existem dois tipos principais de operações

entre investidores:

Negócio Efetivo: Neste tipo de operação, uma financeira vende um título de crédito privado a outra empresa financeira, como a transferência de uma debênture de um fundo de investimento para outro fundo gerido por uma instituição financeira distinta, por exemplo. Esta operação reflete uma mudança efetiva na titularidade do título, sendo crucial para a formação de preços no mercado secundário.

Passagem Entre Fundos: Esta operação ocorre quando um investidor transfere um título privado entre fundos sob sua própria administração, como no caso de um fundo de investimento transferindo um título para outro fundo gerido pela mesma instituição. Embora essa operação não represente uma transação de mercado tradicional, ela deve ser devidamente registrada, impactando as métricas operacionais do mercado.

Ambas as operações, Negócio Efetivo e Passagem Entre Fundos, são registradas na CETIP e conduzidas por uma corretora, que cobra uma taxa de corretagem. É importante destacar que, por ser uma operação menos complexa, a Passagem Entre Fundos geralmente incorre em uma taxa de corretagem menor do que um Negócio Efetivo. Por exemplo, enquanto para Passagem Entre Fundos é cobrado 20 *basis points* do valor do título, para Negócios Efetivos a taxa é de 50 *basis points*.

No setor de brokeragem e gestão de ativos, uma métrica operacional relevante é o *Average Daily Trading Volume* (ADTV), que representa o volume médio diário de negociação de um ativo no mercado. Esta métrica é amplamente utilizada para avaliar a liquidez dos ativos e para a tomada de decisões de investimento, no estudo realizado por (POPESCU; XU, 2023) sobre a gestão de liquidez por fundos mútuos é possível denotar como diferentes estratégias de liquidez impactam o desempenho dos fundos, especialmente em condições de mercado voláteis como o crédito privado. Contudo, o ADTV não faz distinção entre Passagens Entre Fundos e Negócios Efetivos, o que pode introduzir distorções na interpretação dessa métrica, especialmente em mercados onde as Passagens Entre Fundos são comuns.

A correta interpretação do ADTV e outras métricas relacionadas é essencial para a valoração de ativos no mercado de crédito privado. Uma instituição financeira deve avaliar corretamente suas obrigações para evitar desafios de liquidez e garantir uma eficiente planificação de investimentos futuros (JÚNIOR, 2022). Por outro lado, gestores de recursos de terceiros possuem a responsabilidade legal de estabelecer o valor justo dos ativos em suas carteiras, conforme as diretrizes dos órgãos reguladores.

Neste trabalho, exploraremos a aplicação de técnicas de clusterização assim como Mehta et al. (2019) analisaram o uso de algoritmos de clusterização e classificação para a análise de ações, oferecendo insights sobre estratégias de portfólio baseadas em aprendizado de máquina (MEHTA et al., 2019). Iremos abordar especificamente o algoritmo DBSCAN,

para identificar e analisar padrões no volume médio diário de negociação de ativos de crédito privado. A utilização de técnicas de aprendizado não supervisionado, como o clustering aglomerativo, tem se mostrado altamente eficaz na identificação de pares de negociação ao incorporar características das empresas e informações de preço, o que melhora significativamente o desempenho das estratégias de *trading*, conforme demonstrado por (HAN; HE; TOH, 2023). A análise busca distinguir entre operações de mercado que possam impactar de forma diferente a métrica de ADTV, com foco particular nos clusters com muitas operações, que podem indicar operações não representativas ou negócios que não foram à mercado.

A organização estrutural deste trabalho segue a lógica dos objetivos delineados e do contexto econômico apresentado nas seções anteriores. Inicialmente, a sessão 3 oferece uma visão abrangente sobre o contexto econômico e as teorias de mercado relacionadas ao mercado de crédito privado, abordando as características dos títulos de crédito, com ênfase nas debêntures e nos métodos de precificação. Em seguida, o capítulo 4 explora os algoritmos e técnicas relacionadas, justificando a escolha do DBSCAN para a análise das negociações no mercado secundário. Posteriormente, a seção 5 descreve a metodologia desenvolvida e os experimentos realizados, detalhando as etapas do processo e os resultados obtidos a partir da aplicação do algoritmo DBSCAN nas negociações do ativo PEJA11. Por fim, as conclusões gerais do estudo e as implicações dos resultados são discutidas na seção 6, proporcionando uma síntese das descobertas e sugestões para futuros trabalhos na área.

2 Objetivos

2.1 Objetivos Gerais

O principal objetivo deste projeto é analisar e compreender o impacto das operações de *Passagem Entre Fundos* no mercado secundário de renda fixa para crédito privado. Para atingir esse propósito, será desenvolvido um algoritmo de reconhecimento de padrões nas negociações registradas na CETIP, visando identificar as características típicas das operações de *Passagem Entre Fundos*. Esse estudo se concentrará em uma análise detalhada de uma debênture específica, a PEJA11, escolhida devido ao seu alto volume financeiro negociado e à frequência de suas negociações, que a tornam representativa para a análise proposta.

Originalmente, foi considerado o estudo das 50 debêntures com maior volume financeiro negociado, contudo, conforme sugerido pelo Orientador, a redução do número de debêntures analisadas permite um foco mais preciso no problema de clusterização, facilitando a aplicação em larga escala posteriormente. A análise inicial dos dados, referentes ao período de abril de 2023 a abril de 2024, envolveu a avaliação do volume financeiro e da frequência de negociações dos ativos, resultando na seleção da debênture PEJA11 como o ativo mais adequado para o estudo, devido à sua relevância no mercado e à quantidade substancial de dados disponíveis para análise.

2.2 Objetivos Específicos

Para alcançar o objetivo geral delineado acima, este trabalho se propõe a cumprir os seguintes objetivos específicos:

- Desenvolver um algoritmo de reconhecimento de padrões, utilizando técnicas de aprendizado de máquina não supervisionado, em especial o DBSCAN, para identificar operações de *Passagem Entre Fundos* nas negociações registradas.
- Realizar uma análise detalhada da debênture PEJA11, examinando seu comportamento no mercado secundário, com foco no volume financeiro negociado e na frequência das negociações.
- Comparar as características das operações de *Passagem Entre Fundos* com as operações efetivas, identificando como essas influenciam a métrica de *Average Daily Trading Volume* (ADTV) e a percepção de liquidez em mercados específicos (QIN et al., 2023).

- Aplicar técnicas de visualização de dados para apresentar os resultados da análise, destacando padrões de negociação e anomalias detectadas pelo algoritmo.
- Avaliar a eficácia do algoritmo desenvolvido na identificação de *Passagens Entre Fundos* e discutir como essa metodologia pode ser aplicada em larga escala para outros ativos no mercado de crédito privado.

2.3 Resultados Esperados

Ao final deste trabalho, espera-se obter uma compreensão aprofundada sobre o impacto das operações de *Passagem Entre Fundos* no mercado secundário de renda fixa para crédito privado, particularmente no que se refere à sua influência sobre a métrica de *Average Daily Trading Volume* (ADTV) e à percepção de liquidez dos títulos. O algoritmo desenvolvido deverá ser capaz de identificar com precisão essas operações, diferenciando-as das operações efetivas e revelando padrões e anomalias nas negociações da debênture PEJA11.

Além disso, espera-se que as técnicas de visualização aplicadas ofereçam uma representação clara e intuitiva dos dados analisados, facilitando a interpretação dos resultados e a identificação de tendências relevantes. A partir dessa análise, pretende-se avaliar a viabilidade da aplicação dessa metodologia em outros ativos do mercado de crédito privado, contribuindo para uma prática mais informada e precisa na avaliação da liquidez e na precificação desses títulos.

3 Contexto Econômico e Teoria de Mercado

Este capítulo tem como objetivo apresentar o contexto econômico no qual o mercado de crédito privado está inserido, abordando as principais características dos títulos de crédito privado, com ênfase nas debêntures. A análise incluirá a estrutura do mercado primário e secundário, os métodos de precificação desses títulos, e a importância dos diferentes tipos de indexadores utilizados. Além disso, será explorada a relevância das medidas de liquidez no mercado secundário, um fator crucial para a negociação e avaliação desses ativos. A compreensão desses conceitos é fundamental para embasar as análises subsequentes sobre o volume médio diário de negociação e a identificação de padrões no mercado de debêntures.

3.1 Contexto Econômico

No âmbito do mercado financeiro, os títulos de crédito privado desempenham um papel fundamental no financiamento de empresas e projetos. Esses instrumentos, que incluem debêntures, certificados de recebíveis imobiliários (CRI), certificados de recebíveis do agronegócio (CRA), letras financeiras e certificados de depósito bancário (CDB), representam promessas de pagamento de valores futuros emitidas por empresas ou outras entidades. Em troca, os investidores que compram esses títulos fornecem o capital necessário para o crescimento e a expansão dessas organizações.

3.1.1 Títulos do Mercado de Crédito Privado

Os títulos de crédito privado são emitidos no mercado primário, onde as empresas captam recursos diretamente dos investidores. Posteriormente, esses títulos podem ser negociados no mercado secundário, que proporciona liquidez aos investidores, permitindo a compra e venda dos títulos antes do vencimento. O mercado secundário é vital para a eficiência dos mercados financeiros, pois facilita a gestão de portfólios e a realização de lucros ou a mitigação de perdas pelos investidores.

A entrada de um título no mercado secundário ocorre quando o investidor original decide vender o título a outro investidor. Essa transação não envolve o emissor original do título, mas é crucial para a formação contínua de preços e para a avaliação de riscos no mercado. A liquidez do título no mercado secundário pode variar dependendo de fatores como a solvência do emissor, as condições de mercado e a atratividade dos rendimentos oferecidos. Zuppini (2018) propôs o uso de redes neurais e técnicas de clusterização para a precificação de debêntures ilíquidas, uma abordagem inovadora em mercados de baixa

liquidez (ZUPPINI, 2018).

3.1.2 Indexadores Relacionados às Debêntures

As debêntures podem ser emitidas com diferentes tipos de indexadores, que afetam diretamente a forma como os pagamentos de juros e o valor de resgate do título são calculados. Os principais indexadores utilizados nas debêntures brasileiras incluem:

- **Taxa DI (Depósito Interfinanceiro):** Indexador baseado na taxa média dos depósitos interfinanceiros, amplamente utilizado em debêntures pós-fixadas.
- **IPCA (Índice de Preços ao Consumidor Amplo):** Indexador que ajusta o valor nominal da debênture com base na inflação medida pelo IPCA, proporcionando proteção contra a perda de poder aquisitivo.
- **IGP-M (Índice Geral de Preços – Mercado):** Similar ao IPCA, o IGP-M é outro índice de inflação utilizado para ajustar o valor dos títulos.
- **Prefixada:** Em uma debênture prefixada, a taxa de juros é determinada no momento da emissão e permanece fixa ao longo de todo o período de vigência do título. Essa característica oferece previsibilidade para os investidores, que sabem exatamente quanto receberão de juros ao longo do tempo. Para a empresa emissora, a opção por uma emissão prefixada pode ser motivada pela expectativa de que as taxas de juros de mercado aumentarão no futuro. Ao garantir uma taxa fixa, a empresa protege-se contra a alta dos custos de financiamento, embora assuma o risco de pagar mais caro caso as taxas de juros no mercado venham a cair.

Esses indexadores desempenham um papel fundamental na determinação do valor das debêntures e no cálculo dos rendimentos que os investidores receberão ao longo do tempo.

3.1.3 Debêntures e Métodos de Precificação

Dentre os títulos de crédito privado, as debêntures se destacam como um dos mais importantes e amplamente utilizados. Uma debênture é um título de dívida emitido por empresas para captar recursos a longo prazo, com a promessa de pagamento de juros regulares (cupons) e do valor principal no vencimento. As debêntures podem ser emitidas com ou sem garantias, e seus rendimentos podem ser prefixados, pós-fixados ou híbridos.

A precificação de debêntures no mercado secundário é um processo complexo que leva em consideração diversos fatores, como a taxa de juros do mercado, o prazo até o vencimento, a qualidade de crédito do emissor e a liquidez do título. Valle (2020) analisou

a reavaliação de preços de debêntures no mercado secundário brasileiro, fornecendo um estudo de caso detalhado sobre a volatilidade desses títulos (VALLE, 2020).

Ao contrário de outros ativos, as debêntures são geralmente negociadas em termos de taxa de rendimento, e não de preço unitário. Essa prática reflete a relação inversa entre a taxa de rendimento de uma debênture e seu preço de mercado: quando a taxa de rendimento aumenta, o preço do título tende a cair, e vice-versa.

O método de precificação de debêntures adotado pelo mercado envolve o cálculo do valor presente dos fluxos de caixa futuros que o título gerará, descontados pela taxa de rendimento prevalecente. A Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais (ANBIMA) desempenha um papel crucial nesse processo, ao fornecer taxas indicativas para a precificação de debêntures. Essas taxas são baseadas em negócios efetivos registrados no sistema CETIP, calls de corretoras e dados coletados junto aos principais participantes do mercado.

3.1.4 Método de Precificação da ANBIMA

A ANBIMA, como referência no mercado financeiro, oferece uma metodologia robusta para a precificação de debêntures, especialmente no mercado secundário, onde esses títulos são negociados em termos de taxa e não em preço unitário. Esse método considera as características específicas de cada título, como o prazo até o vencimento, a qualidade de crédito do emissor e o tipo de indexador associado.

A precificação em taxa, e não em preço unitário, é particularmente relevante em mercados com alta volatilidade, pois permite que os investidores comparem o rendimento oferecido por diferentes debêntures de forma padronizada. A ANBIMA coleta e consolida informações de diversas fontes, incluindo transações registradas no CETIP e estimativas fornecidas por corretoras e outros participantes do mercado, para divulgar taxas indicativas que auxiliam os investidores na avaliação das debêntures disponíveis.

Essas taxas indicativas, ao serem aplicadas no cálculo do valor presente dos fluxos de caixa esperados, fornecem uma estimativa do valor justo das debêntures no mercado secundário, refletindo as condições de mercado e os riscos associados a cada título.

3.2 Importância das Medidas de Liquidez no Mercado Secundário

A liquidez é um dos pilares fundamentais para a eficiência e funcionamento adequado do mercado secundário. No contexto dos títulos de crédito privado, a liquidez se refere à facilidade com que um título pode ser comprado ou vendido no mercado, tentando ao máximo não causar uma variação significativa em seu preço. Títulos com alta liquidez são atraentes para os investidores, pois oferecem maior flexibilidade na gestão de portfólios,

permitindo que ajustes sejam feitos rapidamente em resposta a mudanças nas condições de mercado ou nas estratégias de investimento. O Artigo de (PHAM; SINGH; VU, 2023) destaca como as notícias podem influenciar rapidamente a percepção de liquidez no mercado.

Medidas de liquidez, como o *Average Daily Trading Volume* (ADTV), são frequentemente utilizadas para avaliar o nível de atividade em torno de um título específico. Um ADTV elevado indica que o título é negociado com frequência, sugerindo que há um mercado ativo e que o título pode ser liquidado rapidamente com pouco impacto sobre o preço. Em contrapartida, títulos com baixo ADTV podem enfrentar desafios de liquidez, sendo mais suscetíveis a variações de preço ao tentar realizar transações de maior volume. Gao e Bai (2023) demonstraram a importância de incorporar o risco de liquidez na precificação de opções financeiras, uma abordagem relevante também para títulos de crédito privado (GAO; BAI, 2023).

A liquidez também desempenha um papel crucial na precificação dos títulos no mercado secundário. Títulos com menor liquidez geralmente exigem um prêmio de risco maior, refletido em taxas de rendimento mais elevadas, para compensar os investidores pela dificuldade potencial de vender o título em momentos de necessidade. Assim, a avaliação da liquidez é essencial tanto para investidores quanto para emissores, pois impacta diretamente o custo de captação de recursos e a rentabilidade das operações.

No contexto do mercado de debêntures, a análise da liquidez é particularmente relevante devido à diversidade de emissores, prazos de vencimento e características de risco. Ao integrar medidas de liquidez na análise de mercado, é possível obter uma visão mais clara sobre a atratividade e o comportamento dos títulos, o que é fundamental para a tomada de decisões de investimento informadas e eficazes. Naik et al. (2020) exploraram as várias dimensões da liquidez no mercado indiano, fornecendo uma base sólida para a análise de liquidez em outros mercados emergentes (NAIK; POORNIMA; REDDY, 2020).

4 Algoritmos e técnicas relacionadas

4.1 Escolha do Algoritmo Utilizado

A escolha do algoritmo para análise de dados é uma etapa crucial em qualquer projeto de ciência de dados. Estudos como o de Cai et al. (2016) revisaram diversas abordagens de clusterização, destacando a relevância dessas técnicas na análise de dados financeiros complexos (CAI; LE-KHAC; KECHADI, 2016). No contexto deste trabalho, onde o objetivo é identificar padrões e outliers no volume médio diário de negociação de debêntures, foi considerada a revisão abrangente das técnicas de clusterização, principalmente na área de aprendizado não supervisionado (SAXENA et al., 2017) mas, optou-se pela utilização do algoritmo DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*).

4.1.1 Motivação para o Uso do DBSCAN

O DBSCAN é um algoritmo de aprendizado de máquina não supervisionado amplamente reconhecido por sua capacidade de identificar clusters de alta densidade em conjuntos de dados, bem como de distinguir pontos de dados que não pertencem a nenhum cluster, denominados outliers. Essa característica é particularmente valiosa no mercado de crédito privado, onde as debêntures podem apresentar variações significativas em termos de liquidez, volume de negociação e risco de crédito. Essas variações muitas vezes resultam em padrões complexos que não são facilmente capturados por métodos tradicionais.

Uma das principais vantagens do DBSCAN é que ele não requer a especificação prévia do número de clusters, ao contrário de algoritmos como o K-means, que, apesar de seu bom desempenho, pode ser menos flexível em cenários com dados financeiros variáveis (BINI; MATHEW, 2016). Essa flexibilidade é especialmente relevante no contexto em questão, onde a estrutura dos dados financeiros pode ser altamente variável e onde o número de clusters não é conhecido de antemão. Além disso, o DBSCAN é eficaz na identificação de clusters de formas arbitrárias, o que é essencial ao lidar com dados financeiros complexos, como os volumes de negociação de debêntures, que podem não seguir padrões de distribuição simples.

O DBSCAN também se destaca por sua robustez no tratamento de ruídos e outliers, o que é uma vantagem significativa na análise de dados de mercado, que frequentemente incluem transações atípicas ou anômalas. Em vez de distorcer os resultados, o algoritmo identifica e isola essas anomalias, permitindo uma análise mais precisa dos clusters principais.

4.1.2 Considerações sobre a Não Utilização de Técnicas Supervisionadas

A opção por não utilizar algoritmos de aprendizado supervisionado, como o K-Nearest Neighbors (KNN), baseou-se em várias considerações relacionadas à natureza dos dados e aos objetivos deste estudo. O KNN, por exemplo, é um algoritmo que requer um conjunto de dados rotulado para o treinamento do modelo, o que significa que seria necessário dispor de exemplos previamente classificados para cada título de debênture.

No entanto, no mercado de debêntures, cada título pode apresentar características únicas, como o risco de crédito específico da emissão, a maturidade, o tipo de indexador e a liquidez. Treinar um modelo com base em dados de apenas um título ou de um conjunto limitado de títulos poderia introduzir vieses consideráveis, comprometendo a generalização do modelo para outros títulos com características distintas. Além disso, a obtenção de um volume substancial de dados rotulados necessário para treinar um modelo supervisionado eficaz pode ser inviável, especialmente em mercados financeiros dinâmicos.

Das et al. (2024) mostrou a eficácia da clusterização baseada em densidade quando aplicada na análise de mercados financeiros, especialmente em ambientes de alta volatilidade (DAS; SAHA, 2024). Portanto, diante dessas considerações, a escolha recaiu sobre um algoritmo de aprendizado não supervisionado como o DBSCAN, que não exige rótulos de dados e é capaz de identificar padrões e anomalias diretamente dos dados brutos. Essa abordagem proporciona uma análise mais flexível e robusta, adaptando-se melhor às nuances e variações inerentes ao mercado de crédito privado.

Em síntese, a escolha do DBSCAN fundamentou-se em sua capacidade de lidar com a complexidade e a variabilidade dos dados de negociação de debêntures, sem a necessidade de um extenso conjunto de dados rotulados. Essa decisão reflete uma abordagem prática e eficaz para a análise exploratória de dados financeiros, permitindo a descoberta de padrões significativos que podem informar decisões de investimento.

4.2 Uso do DBSCAN e Escolha do Parâmetro Eps

A aplicação do DBSCAN requer a definição de dois parâmetros principais: eps (epsilon), que define o raio máximo de proximidade entre dois pontos para que sejam considerados vizinhos, e $minPts$, que determina o número mínimo de pontos necessários para formar um cluster. A escolha adequada desses parâmetros é crucial para o desempenho do algoritmo e a qualidade dos clusters identificados.

4.2.1 Escolha do Parâmetro Eps

O parâmetro eps desempenha um papel fundamental no DBSCAN, pois determina o tamanho dos clusters formados e a sensibilidade do algoritmo a outliers. A escolha do eps

deve ser feita de forma cuidadosa, considerando as características específicas dos dados.

Uma prática comum na escolha do *eps* é a utilização do gráfico *k-distance*, onde as distâncias dos pontos ao seu *k*-ésimo vizinho mais próximo são ordenadas e plotadas. A escolha do *eps* é geralmente feita na região onde o gráfico exibe uma curva acentuada, indicando uma transição entre as distâncias intra-cluster e inter-cluster. No contexto deste trabalho, o valor de *eps* foi ajustado de acordo com a distribuição das taxas de negociação da debênture PEJA11, utilizando o gráfico *k-distance* como referência para identificar o ponto de inflexão mais apropriado.

4.2.2 Justificativa para a Escolha de *minPts* igual a 2

A escolha do parâmetro *minPts*, que define o número mínimo de pontos necessários para formar um cluster no algoritmo DBSCAN, é uma decisão importante, pois influencia diretamente a definição dos clusters e a detecção de outliers. Neste trabalho, adotou-se o valor de *minPts* igual a 2, o que se justifica devido à natureza das operações de *Passagens Entre Fundos*.

No mercado de crédito privado, as operações de *Passagens Entre Fundos* sempre envolvem, no mínimo, duas entidades: uma parte que vende o título e outra que compra. Assim, cada transação é constituída por um par de entidades envolvidas no processo de transferência de ativos. Portanto, o valor de *minPts* igual a 2 reflete essa característica intrínseca, assegurando que o DBSCAN considere essas operações como clusters válidos apenas quando há pelo menos um par de transações, ou seja, uma compra e uma venda.

4.2.3 Adaptação do DBSCAN às Características dos Dados

A adaptabilidade do DBSCAN é uma de suas principais vantagens, especialmente em cenários onde os dados podem apresentar densidades variáveis e estruturas complexas. No mercado de crédito privado, onde as debêntures podem ter padrões de negociação influenciados por fatores como risco de crédito, maturidade e condições macroeconômicas, o DBSCAN se mostrou eficaz na separação de clusters significativos e na identificação de outliers, que podem representar operações atípicas ou *Passagens entre Fundos*.

Estudos recentes têm demonstrado a eficácia do DBSCAN em diversos domínios, incluindo o mercado financeiro. Por exemplo, a abordagem proposta por Dinandra et al. (2019), que combina algoritmos genéticos com clusterização, foi utilizada para otimizar portfólios, mostrando retornos superiores (DINANDRA; HERTONO; HANDARI, 2019). Além disso, Dubey et al. (2018) exploraram o uso de análise de clusterização para detecção de outliers em títulos de renda fixa, contribuindo para a gestão de risco em mercados financeiros (DUBEY et al., 2018).

5 Metodologia Desenvolvida e Experimentos

Conforme mencionado na Introdução (1), este trabalho faz parte do objetivo geral do sistema e marca o início do projeto. Este capítulo descreve detalhadamente a metodologia desenvolvida para a análise das operações de *Passagens entre Fundos* no mercado secundário de debêntures, com foco na identificação e diferenciação dessas operações das negociações efetivas. O capítulo é dividido em duas partes principais: a primeira aborda os métodos aplicados, enquanto a segunda discute os resultados experimentais obtidos a partir desses métodos.

5.1 Métodos

Nesta seção, detalhamos as etapas do processo metodológico utilizado para alcançar os objetivos propostos. A seguir, apresentamos os principais passos envolvidos na metodologia:

1. **Extração dos dados:** Os dados são extraídos das coletas diárias realizadas na CETIP, contendo informações detalhadas sobre as negociações de debêntures, incluindo volume financeiro, quantidade negociada, taxas de negócio, entre outros parâmetros relevantes para o estudo. A coleta foi realizada de maneira automatizada, abrangendo o período de abril de 2023 a abril de 2024.
2. **Tratamento dos dados:** Após a extração, os dados passaram por um rigoroso processo de tratamento para garantir que apenas informações completas e relevantes fossem utilizadas nas análises subsequentes. Dados com inconsistências, como registros incompletos ou ausência de informações essenciais, foram descartados ou corrigidos quando possível. Esse processo de limpeza foi fundamental para assegurar a precisão e a confiabilidade dos resultados.
3. **Separação de debêntures e seleção do ativo PEJA11:** Com os dados devidamente tratados, foi realizada a separação dos registros de acordo com as debêntures negociadas. O ativo PEJA11 foi escolhido como foco principal da análise, devido ao seu alto volume financeiro e à frequência significativa de negociações durante o período analisado. A análise que motivou a seleção do PEJA11 é apresentada nas figuras 1 e 2.
4. **Agrupamento por data de liquidação e quantidade:** Para uma análise mais detalhada dos padrões de negociação, as transações do ativo PEJA11 foram agrupadas por data de liquidação e quantidade negociada. Esse agrupamento permite observar

as transações com base em sua temporalidade e magnitude, facilitando a identificação de padrões recorrentes, essenciais para a aplicação do algoritmo DBSCAN.

5. **Aplicação do algoritmo DBSCAN:** O DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) foi escolhido para a clusterização das negociações. Esse algoritmo é especialmente útil para identificar clusters de alta densidade e outliers, características essenciais para separar operações de *Passagens entre Fundos* das negociações efetivas. A aplicação do DBSCAN nas transações do PEJA11 é ilustrada na Figura 4.
6. **Remoção de falsos positivos e refinamento dos clusters:** Após a aplicação do DBSCAN, foi necessário realizar um refinamento dos clusters, removendo falsos positivos e classificações que não correspondiam a *Passagens entre Fundos*. Esse refinamento aumentou a precisão dos resultados, garantindo que apenas os padrões de interesse fossem mantidos para a análise subsequente. O processo de refinamento é demonstrado na Figura 6.
7. **Análise comparativa dos resultados:** Por fim, os resultados obtidos após a aplicação do DBSCAN e o refinamento dos clusters foram comparados com os dados originais, sem o uso do algoritmo. Essa análise comparativa revelou uma redução significativa de aproximadamente 28% no *Average Daily Trading Volume (ADTV)*, destacando o impacto das operações de *Passagens entre Fundos* na percepção de liquidez do mercado. A Figura 7 mostra a comparação dos resultados antes e depois da aplicação do DBSCAN.

5.1.1 Extração dos Dados

A primeira etapa do processo metodológico consiste na extração dos dados de negociação de debêntures, obtidos a partir de coletas diárias realizadas na CETIP, disponíveis em <https://tinyurl.com/DadosCetip>. Esses dados incluem informações sobre os negócios efetuados, tais como volume financeiro, quantidade negociada, taxas de negócio, entre outros parâmetros relevantes para o estudo. A coleta foi realizada abrangendo o período de abril de 2023 a abril de 2024, de maneira manual e diária, dado que a informação disponível continha apenas os últimos 20 dias úteis.

5.1.2 Tratamento dos Dados

Após a extração, os dados passaram por um rigoroso processo de tratamento. Esta etapa visa garantir que apenas informações completas e relevantes sejam utilizadas nas análises subsequentes. Dados com inconsistências, como registros cancelados ou ausência de informações sobre taxa, foram descartados ou corrigidos quando possível. Esse processo

de limpeza é crucial para assegurar a precisão e a confiabilidade dos resultados obtidos nas etapas posteriores.

5.1.3 Separação de Debêntures e Seleção do Ativo PEJA11

Com os dados devidamente tratados, foi realizada a separação dos registros de acordo com as debêntures negociadas. O ativo PEJA11 (*Debênture pertencente à empresa PETRO RIO JAGUAR PETROLEO S.A. com emissão em Agosto de 2022 e classificada no setor de Petróleo e Gás*) foi escolhido como foco principal da análise, devido ao seu alto volume financeiro operado e à frequência significativa de negociações durante o período analisado. As Figuras 1 e 2 apresentam a análise que motivou a seleção do PEJA11 como ativo representativo para este estudo, uma vez que o ativo está no Top 10 de volume negociado do período, é o segundo com maior número de negociações e não apresenta data de emissão ou vencimento concomitante com o período analisado, como no caso de IRJS14 que teve sua emissão em Maio de 2023.

Codigo IF	Emissor	Quantidade Negociada	Preco Negocio	Volume Financeiro R\$	Taxa Negocio	Horario Negocio	Data Negocio	Data Liquidacao
IRJS14	22223	22223	22223	22223	22223	22223	22223	22223
PEJA11	17797	17797	17797	17797	17797	17797	17797	17797
ENAT11	16875	16875	16875	16875	16875	16875	16875	16875
RMSA12	12116	12116	12116	12116	12116	12116	12116	12116
AEAB11	11828	11828	11828	11828	11828	11828	11828	11828
MTRJ19	11659	11659	11659	11659	11659	11659	11659	11659
ORIG21	11625	11625	11625	11625	11625	11625	11625	11625
ELET14	10994	10994	10994	10994	10994	10994	10994	10994
PLSB1A	10987	10987	10987	10987	10987	10987	10987	10987
ENEV29	10574	10574	10574	10574	10574	10574	10574	10574

Figura 1 – Os 10 ativos mais negociados no período.

```

Codigo IF
RAHD11    222.717.010.387,41
TBSP11    7.089.388.238,77
CELG12    5.506.212.688,62
ELET14    5.184.986.765,43
CESE11    4.931.195.549,73
IRJS14    4.896.059.391,43
TOTS14    3.340.016.196,34
PEJA11    3.183.905.130,72
NIGP12    3.173.074.724,70
CGEE11    3.157.182.807,48
Name: Volume Financeiro R$, dtype: object

```

Figura 2 – As 10 maiores movimentações financeiras de abril de 2023 a abril de 2024.

5.1.4 Agrupamento por Data de Liquidação e Quantidade

Para uma análise mais detalhada dos padrões de negociação, as transações do ativo PEJA11 foram agrupadas por data de liquidação e quantidade negociada. Esse

agrupamento permite observar as transações com base em sua temporalidade e magnitude, facilitando a identificação de padrões recorrentes. Essas informações são essenciais para a etapa seguinte, onde o algoritmo DBSCAN é aplicado.

5.1.5 Aplicação do Algoritmo DBSCAN

O DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) foi escolhido para a clusterização das negociações, conforme discutido no capítulo 4. Este algoritmo é especialmente eficaz na identificação de clusters de alta densidade e na detecção de outliers, características cruciais para distinguir operações de *Passagens entre Fundos* das negociações efetivas.

A escolha do parâmetro *eps*, que define o raio máximo para considerar dois pontos como parte do mesmo cluster, foi realizada de forma cuidadosa. Utilizando o gráfico *k-distance*, foi possível identificar o ponto de inflexão onde a transição entre distâncias intra-cluster e inter-cluster se torna mais pronunciada. Esse ponto de inflexão foi então utilizado para determinar o valor mais apropriado de *eps*.

Para ilustrar o processo de escolha do *eps*, a Figura 3 apresenta uma análise visual dos 12 clusters com o maior número de operações, demonstrando como o *eps* influencia a formação dos clusters. Após cuidadosa avaliação dos gráficos, foi decidido fixar o valor do *eps* em 0.1, oferecendo uma margem que evita o *overfitting*, mas que ainda captura com precisão os padrões de densidade presentes nas transações da debênture PEJA11. Essa escolha garantiu que o DBSCAN separasse de maneira eficaz as operações típicas das *Passagens entre Fundos*.

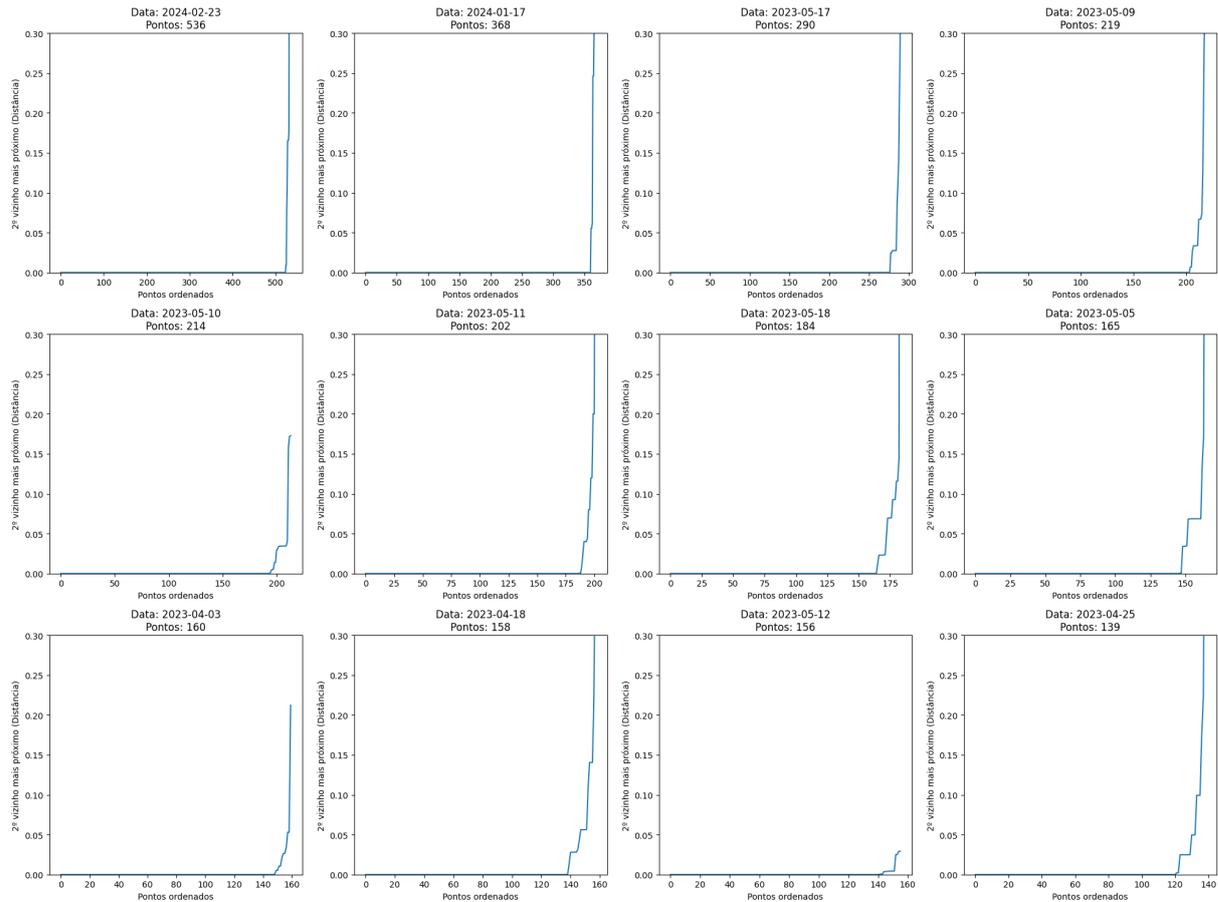


Figura 3 – Análise dos dados para a escolha do valor *eps* com base nos 12 clusters com maior número de operações.

A Figura 4 ilustra a distribuição das taxas de negociação por cluster, conforme definido pelo DBSCAN nas transações de PEJA11. Um ponto de destaque é a alta frequência de operações ocorrendo na mesma taxa, um indicativo de um fenômeno conhecido como *negociação interna*. Esse padrão surge quando grandes conglomerados repassam o volume financeiro institucional para instituições menores, que possuem vínculo jurídico com o mesmo grupo e atendem diretamente aos clientes finais.

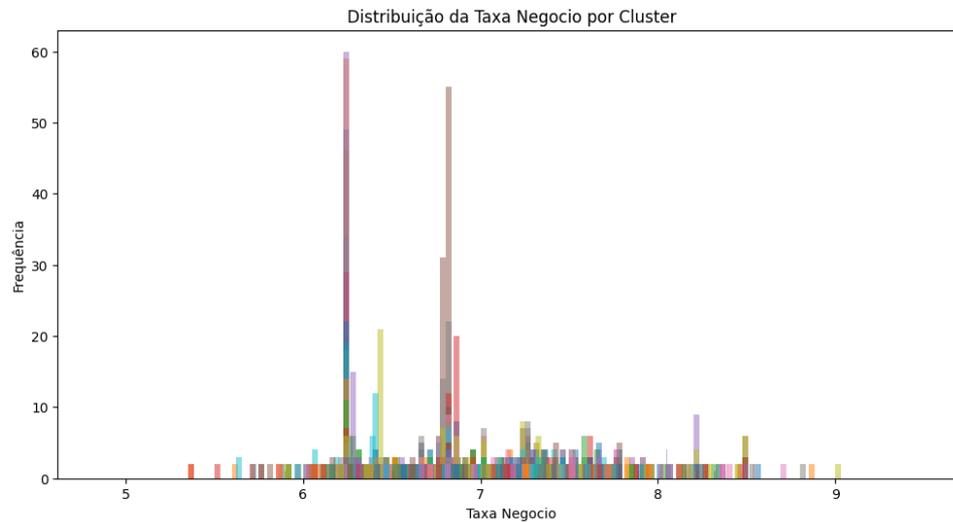


Figura 4 – Aplicação do DBSCAN nas negociações do ativo PEJA11 com base na taxa de negócio.

As operações de *negociação interna*, como demonstrado na Figura 5, apresentam uma contagem elevada de negócios em clusters específicos, influenciando diretamente a média de 2.54 negócios por cluster. Esse tipo de operação, que ocorre predominantemente dentro de conglomerados financeiros, envolve a transferência de títulos entre instituições ligadas ao mesmo grupo, sem que esses negócios sejam efetivamente expostos ao mercado. Como resultado, essas transações podem distorcer a medida de liquidez do *Average Daily Trading Volume* (ADTV), dando uma falsa impressão de alta liquidez. Portanto, para obter uma análise mais precisa da liquidez real no mercado secundário, é essencial identificar e remover essas operações internas. Ao fazer isso, podemos evitar que elas contaminem as métricas de liquidez, proporcionando uma visão mais clara e precisa do comportamento de mercado para os investidores e reguladores.

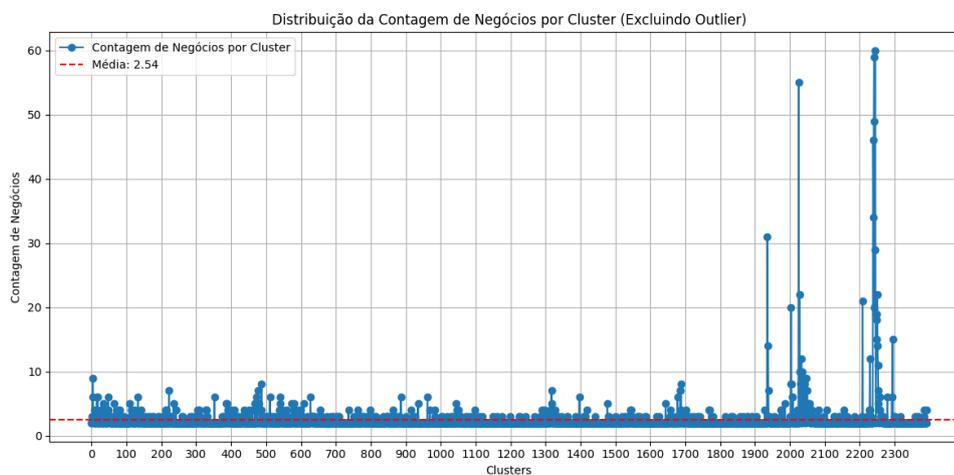


Figura 5 – Distribuição da Contagem de Negócios por Clusters que não são Outliers.

5.1.6 Remoção de Falsos Positivos e Refinamento dos Clusters

Após a aplicação do DBSCAN, foi necessário realizar um refinamento dos clusters, removendo falsos positivos e classificações que não correspondiam a *Passagens entre Fundos*. Esse refinamento aumenta a precisão dos resultados, garantindo que apenas os padrões de interesse sejam mantidos para a análise subsequente. Em particular, as operações de *negociação interna*, que tendem a distorcer as métricas de liquidez ao criar uma falsa percepção de atividade de mercado, foram identificadas e filtradas durante este processo. Como essas operações geralmente não são expostas ao mercado e são realizadas entre entidades do mesmo conglomerado, elas podem inflar artificialmente o *Average Daily Trading Volume* (ADTV) e outros indicadores de liquidez. Portanto, a remoção dessas transações é crucial para assegurar que a análise reflète com precisão as verdadeiras dinâmicas do mercado. A Figura 6 demonstra o processo de refinamento realizado, incluindo a eliminação dessas operações para uma análise mais fidedigna.

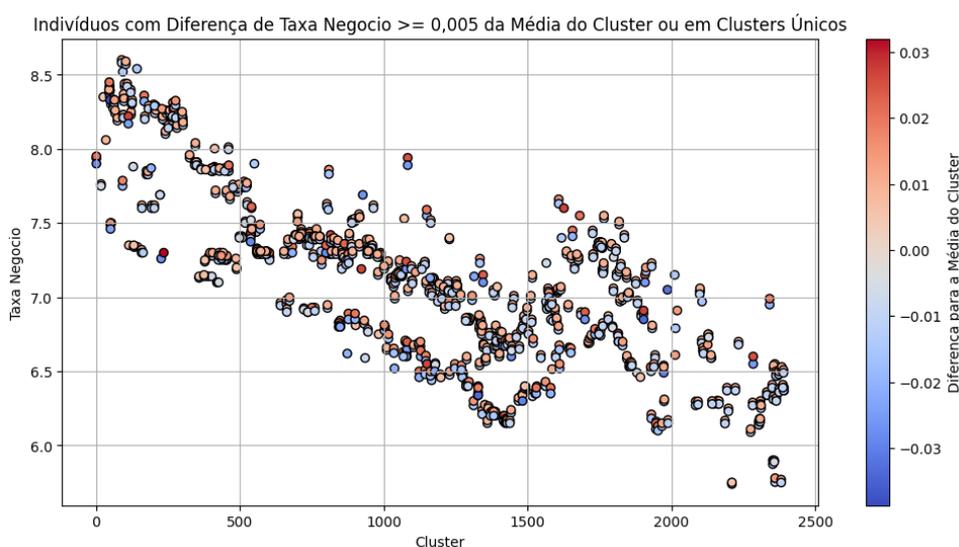


Figura 6 – Refinamento dos clusters após a aplicação do DBSCAN, com remoção de falsos positivos.

5.2 Resultados Experimentais

Nesta seção, discutiremos os resultados experimentais obtidos a partir da aplicação da metodologia desenvolvida, com foco na análise das métricas de liquidez e na identificação de padrões de negociação no mercado de debêntures após a aplicação do DBSCAN. A influência das operações de *Passagens entre Fundos* é também explorada, especialmente no que diz respeito ao impacto nas taxas de negociação intradiárias e no *Average Daily Trading Volume* (ADTV).

5.2.1 Análise Comparativa das Métricas de Liquidez

Após a aplicação do DBSCAN e o subsequente refinamento dos clusters, realizamos uma análise comparativa das principais métricas de liquidez antes e depois do processamento. As métricas analisadas incluem o *Average Daily Trading Volume* (ADTV), o número de negócios diários e a volatilidade das taxas de negociação. A Tabela 1 apresenta uma síntese dessas comparações.

Tabela 1 – Comparação das Métricas de Liquidez antes e depois da Aplicação do Algoritmo

Métrica	ADTV (milhões de R\$)	Número Médio de Negócios Diários
Antes do Algoritmo	12,57	70,32
Depois do Algoritmo	9,05	50,79
Diferença (% - Módulo)	27,99%	27,77%

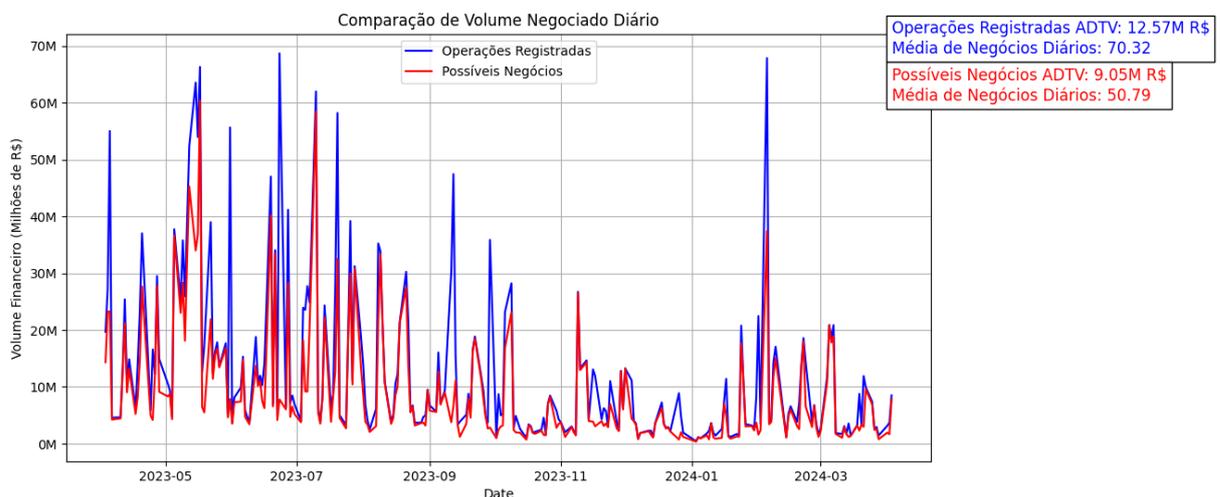


Figura 7 – Análise comparativa dos resultados antes e após a aplicação do Algoritmo.

Os resultados mostram uma redução significativa no ADTV, de aproximadamente 28%, após a remoção das operações de *Passagens entre Fundos*. O número médio de negócios diários também diminuiu, refletindo a exclusão de transações que não contribuem para a liquidez real do mercado. A volatilidade das taxas de negociação foi ligeiramente reduzida, sugerindo uma estabilização do mercado após a eliminação das operações internas.

5.2.2 Influência das *Passagens entre Fundos* nas Taxas de Negociação Intradiaárias

Além das métricas de liquidez, analisamos a influência das *Passagens entre Fundos* nas taxas de negociação intradiárias do mercado secundário. Assim como (BYRD; BALCH, 2019) exploraram o uso de aprendizado profundo para prever preços intradiários, oferecendo insights valiosos sobre a eficiência de mercado, a hipótese era de que estas operações tendem

a criar artificialmente uma alta frequência de negociações em taxas específicas, distorcendo a percepção de atividade de mercado.

A Figura 8 apresenta uma comparação das taxas de negociação intradiárias antes e depois da aplicação do algoritmo. Como pode ser observado, a presença das *Passagens entre Fundos* resulta em picos de atividade em determinadas faixas de taxa, que desaparecem ou são significativamente reduzidos após a remoção dessas operações. Porém o impacto a longo prazo é dissipado ou mesmo contido, podendo ter uma variação brusca diária, mas tendo pouco efeito no processo de precificação da Anbima, por exemplo, dado que ele não precifica dependendo apenas do dia anterior, mas também com base no histórico de preços gerados.

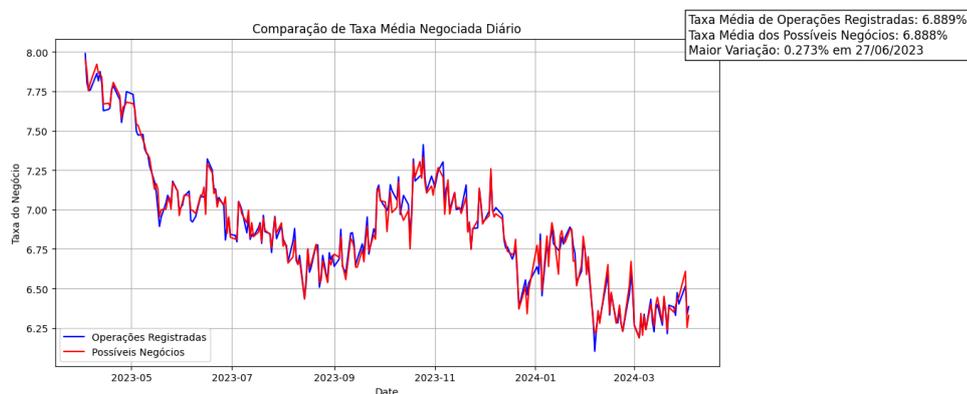


Figura 8 – Comparação das Taxas de Negociação Intradiárias antes e depois da Aplicação do Algoritmo.

Essa análise demonstra como as *Passagens entre Fundos* podem afetar a integridade das métricas de mercado, criando uma falsa impressão de liquidez e distorcendo a verdadeira dinâmica de oferta e demanda. A remoção dessas operações, portanto, é crucial para uma avaliação mais precisa do comportamento de mercado.

5.3 Discussão dos Resultados

Os resultados obtidos a partir da aplicação do algoritmo DBSCAN e do refinamento dos clusters mostraram-se altamente eficazes na identificação e separação das operações de *Passagens entre Fundos* das negociações efetivas no mercado secundário de debêntures.

A análise comparativa entre os dados processados e os dados originais, conforme ilustrado na Figura 7, revelou uma redução significativa de aproximadamente 28% no *Average Daily Trading Volume* (ADTV). Essa redução destaca a influência considerável das *Passagens entre Fundos* na percepção de liquidez do mercado, demonstrando que a métrica de ADTV, sem o devido tratamento, pode ser inflada artificialmente por operações que não representam uma verdadeira negociação de mercado.

Além disso, a Figura 8 evidencia o impacto dessas operações nas taxas de negociação intradiárias. Observa-se que as *Passagens entre Fundos* geram picos de atividade em determinadas faixas de taxa, que são suprimidos ou significativamente reduzidos após a remoção dessas operações pelo algoritmo. Embora esses picos possam indicar uma elevada atividade em curto prazo, eles não refletem uma mudança real no valor de mercado dos títulos, mas sim uma distorção causada por negociações internas dentro de conglomerados financeiros.

Entretanto, é importante notar que, a longo prazo, o impacto dessas operações tende a ser dissipado. Isso se deve ao fato de que o processo de precificação, como o realizado pela ANBIMA, não depende exclusivamente das taxas de negociação do dia anterior, mas também de uma análise histórica mais abrangente. Dessa forma, a influência das *Passagens entre Fundos* sobre a precificação diária é mitigada pela consideração de um conjunto mais amplo de dados, o que assegura que a precificação reflète de maneira mais precisa o valor real de mercado dos ativos.

Esses resultados sublinham a eficácia da metodologia aplicada e sugerem que a incorporação de técnicas de aprendizado de máquina, como o DBSCAN, pode melhorar substancialmente a qualidade das análises de liquidez e precificação no mercado de crédito privado. Em última análise, isso pode contribuir para decisões de investimento mais informadas e para uma melhor gestão dos riscos associados a essas operações.

6 Conclusões Finais

Neste trabalho, exploramos a aplicação do algoritmo de aprendizado de máquina não supervisionado DBSCAN para identificar e analisar as operações de *Passagem Entre Fundos* no mercado secundário de renda fixa para crédito privado, com foco na debênture PEJA11. O principal objetivo foi compreender o impacto dessas operações na métrica de *Average Daily Trading Volume* (ADTV) e na percepção de liquidez no mercado.

Os resultados obtidos demonstraram que o algoritmo DBSCAN foi eficaz na identificação dessas operações, permitindo a diferenciação clara entre negociações efetivas e *Passagens Entre Fundos*. Após a aplicação do algoritmo, observou-se uma redução significativa de aproximadamente 28% no ADTV, o que destaca a influência considerável dessas operações na métrica de liquidez do mercado. Além disso, a análise das taxas de negociação intradiárias evidenciou que as *Passagens Entre Fundos* tendem a criar picos de atividade em determinadas faixas de taxa, distorcendo a verdadeira dinâmica de oferta e demanda no mercado secundário.

A remoção dessas operações revelou-se essencial para proporcionar uma avaliação mais precisa da liquidez real dos ativos, auxiliando tanto investidores quanto reguladores a tomar decisões mais informadas. A metodologia desenvolvida neste trabalho pode, portanto, ser aplicada a outros ativos no mercado de crédito privado, contribuindo para uma melhor compreensão das dinâmicas de liquidez e precificação desses títulos.

6.1 Trabalhos Futuros

Embora os resultados obtidos neste estudo tenham sido significativos, há diversas direções futuras que podem ser exploradas para ampliar o escopo e a aplicabilidade da metodologia desenvolvida.

Uma das principais sugestões para trabalhos futuros é a aplicação do algoritmo DBSCAN em ativos considerados ilíquidos no mercado de crédito privado. A hipótese é que, em ativos com menor volume de negociação, as *Passagens Entre Fundos* podem ter um impacto ainda mais pronunciado sobre as métricas de liquidez e precificação. Esses ativos, devido à sua baixa liquidez, podem ser mais sensíveis a operações internas, resultando em distorções ainda maiores nas medidas de ADTV e volatilidade das taxas.

Além disso, uma análise comparativa entre ativos líquidos e ilíquidos pode fornecer insights valiosos sobre como diferentes níveis de liquidez afetam a robustez e a eficácia do algoritmo na identificação de operações atípicas. Esse estudo pode contribuir para o desenvolvimento de estratégias de mitigação de riscos em mercados menos líquidos, onde

as distorções causadas por operações internas são mais difíceis de detectar e corrigir.

Outro possível direcionamento é a adaptação do DBSCAN para trabalhar com dados temporais mais complexos, permitindo a detecção de padrões em séries temporais longas e a análise de impactos de eventos externos sobre a liquidez e precificação de ativos. Isso incluiria o uso de técnicas avançadas de machine learning para prever o comportamento futuro do mercado com base nos padrões identificados.

Finalmente, a integração da metodologia com outras técnicas de análise financeira, como o uso de redes neurais para a previsão de preços ou a aplicação de algoritmos de otimização de portfólios, pode expandir as capacidades analíticas e práticas da abordagem aqui proposta, fornecendo um conjunto de ferramentas mais completo para gestores de ativos e analistas financeiros.

Em conclusão, este trabalho representa um passo significativo na compreensão das dinâmicas de liquidez e precificação no mercado de crédito privado, mas também abre várias novas possibilidades para a aplicação de técnicas de aprendizado de máquina em finanças, especialmente em mercados menos líquidos e mais suscetíveis a distorções operacionais.

Referências

- BINI, B.; MATHEW, T. Clustering and regression techniques for stock prediction. *Procedia Technology*, v. 24, p. 1248–1255, 2016. ISSN 2212-0173. International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015). Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2212017316301931>>. Citado na página 20.
- BYRD, D.; BALCH, T. H. Intra-day equity price prediction using deep learning as a measure of market efficiency. *arXiv preprint arXiv:1908.08168*, 2019. Citado na página 30.
- CAI, F.; LE-KHAC, N.-A.; KECHADI, T. *Clustering Approaches for Financial Data Analysis: a Survey*. 2016. Disponível em: <<https://arxiv.org/abs/1609.08520>>. Citado na página 20.
- DAS, A. H. T.; SAHA, G. Application of density-based clustering approaches for stock market analysis. *Applied Artificial Intelligence*, Taylor & Francis, v. 38, n. 1, p. 2321550, 2024. Disponível em: <<https://doi.org/10.1080/08839514.2024.2321550>>. Citado na página 21.
- DINANDRA, R. S.; HERTONO, G. F.; HANDARI, B. D. Implementation of density-based spatial clustering of application with noise and genetic algorithm in portfolio optimization with constraint. *AIP Conference Proceedings*, v. 2168, n. 1, p. 020026, 11 2019. ISSN 0094-243X. Disponível em: <<https://doi.org/10.1063/1.5132453>>. Citado na página 22.
- DUBEY, G. et al. Outlier detection using cluster analysis for fixed income bonds. In: *2018 8th International Conference on Cloud Computing, Data Science Engineering (Confluence)*. [S.l.: s.n.], 2018. p. 14–15. Citado na página 22.
- GAO, R.; BAI, Y. Pricing quanto options with market liquidity risk. *PLOS ONE*, Public Library of Science, v. 18, n. 9, p. 1–21, 09 2023. Disponível em: <<https://doi.org/10.1371/journal.pone.0292324>>. Citado na página 19.
- HAN, C.; HE, Z.; TOH, A. J. W. Pairs trading via unsupervised learning. *European Journal of Operational Research*, v. 307, n. 2, p. 929–947, 2023. ISSN 0377-2217. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S037722172200769X>>. Citado na página 13.
- JÚNIOR, J. B. de A. CPC 46 - Mensuração do valor justo - técnicas de avaliação. *REVISTA PROCESSUS DE ESTUDOS DE GESTÃO, JURÍDICOS E FINANCEIROS*, v. 12, n. 42, mar. 2022. Disponível em: <<https://doi.org/10.5281/zenodo.6385291>>. Citado na página 12.
- MEHTA, K. et al. Analysis of stocks by the use of clustering and classification algorithms. In: *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*. [S.l.: s.n.], 2019. p. 1–5. Citado na página 12.
- NAIK, P.; POORNIMA, B. G.; REDDY, Y. V. Measuring liquidity in indian stock market: A dimensional perspective. *PLOS ONE*, Public Library of Science, v. 15, n. 9, p.

- 1–17, 09 2020. Disponível em: <<https://doi.org/10.1371/journal.pone.0238718>>. Citado na página 19.
- PHAM, T. P.; SINGH, H.; VU, V. The impact of bank loan announcements on stock liquidity. *SSRN Electronic Journal*, 2023. Citado na página 19.
- POPESCU, M.; XU, Z. Liquidity management by mutual funds. *Managerial Finance*, 2023. Citado na página 12.
- QIN, Z. et al. Will the negative psychological perceptions of investors reduce platform liquidity? evidence from china's online loans. *PLOS ONE*, v. 18, 2023. Citado na página 14.
- SAXENA, A. et al. A review of clustering techniques and developments. *Neurocomputing*, v. 267, p. 664–681, 2017. ISSN 0925-2312. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231217311815>>. Citado na página 20.
- VALLE, M. H. Estudo de caso: a reavaliação dos preços das debêntures no mercado secundário brasileiro durante o 4º trimestre de 2019. <https://bdm.unb.br/handle/10483/29056>, 2020. Citado na página 18.
- ZUPPINI, M. S. *Apuração de debêntures ilíquidas utilizando redes neurais e clustering*. Tese (Doutorado) — FGV, 2018. Citado na página 17.